Measuring precise radial velocities and cross-correlation function line-profile variations using a Skew Normal density[*]

U. Simola[1] [**], X. Dumusque[2] [***], and Jessi Cisewski-Kehe[3]

[1] Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland
[2] Observatoire de Genève, Université de Genève, 51 ch. des Maillettes, CH-1290 Versoix, Switzerland
[3] Department of Statistics and Data Science, Yale University, New Haven, CT, USA

ABSTRACT

Context. Stellar activity is one of the primary limitations to the detection of low-mass exoplanets using the radial-velocity (RV) technique. Stellar activity can be probed by measuring time-dependent variations in the shape of the cross-correlation function (CCF). It is therefore critical to measure with high-precision these shape variations to de-correlate the signal of an exoplanet from spurious RV signals caused by stellar activity.
Aims. We propose to estimate the variations in shape of the CCF by fitting a Skew Normal (SN) density which, unlike the commonly employed Normal density, includes a skewness parameter to capture the asymmetry of the CCF induced by stellar activity and the convective blueshift.
Methods. The performances of the proposed method are compared to the commonly employed Normal density using both simulations and real observations, with different levels of activity and signal-to-noise ratio (SNR) levels.
Results. When considering real observations, the correlation between the RV and the asymmetry of the CCF and between the RV and the width of the CCF are stronger when using the parameters estimated with the SN density rather than the ones obtained with the commonly employed Normal density. In particular the strongest correlations have been obtained when using the mean of the SN as an estimate for the RV. This suggests that the CCF parameters estimated using a SN density are more sensitive to stellar activity, which can be helpful when estimating stellar rotational periods and more in general when characterizing stellar activity signals. Using the proposed SN approach, the uncertainties estimated on the RV defined as the median of the SN are on average 10% smaller than the uncertainties calculated on the mean of the Normal. The uncertainties estimated on the asymmetry parameter of the SN are on average 15% smaller than the uncertainties measured on the Bisector Inverse Slope Span (BIS SPAN), which is the commonly used parameter to evaluate the asymmetry of the CCF.
Conclusions. The SN density is a better model than the Normal density for characterizing the CCF since the correlations used to probe stellar activity are stronger and the uncertainties of the RV estimate and the asymmetry of the CCF are both smaller.

Key words. techniques: radial velocities – planetary systems – stars: activity – methods: data analysis

1. Introduction

When working with radial-velocities data (RVs), one of the main limitations to the detection of low-mass exoplanets is no longer the precision of the instruments used, but the different sources of variability induced by the stars (e.g. Feng et al. 2017; Dumusque et al. 2017; Rajpaul et al. 2015; Robertson et al. 2014). Stellar oscillations, granulation phenomena, and stellar activity can all induce apparent RV signals that are above the meter-per-second ($\mathrm{m\,s^{-1}}$) precision (e.g. Saar & Donahue 1997; Queloz et al. 2001; Desort et al. 2007; Dumusque et al. 2011; Dumusque 2016) reached by the best high-resolution spectrographs (HARPS, HARPS-N, Mayor et al. 2003; Cosentino et al. 2012). It is therefore mandatory to better understand stellar signals and to develop methods to correct for them, if in the near future we want to detect or confirm an Earth-twin planet using the RV technique. This is even more true now that instruments like the Echelle SPectrograph for Rocky Exoplanet and Stable

Spectroscopic Observations (ESPRESSO) (Pepe et al. 2014) and the EXtreme PREcision Spectrometer (EXPRES) (Fischer et al. 2016) should reach the precision and stability to detect such signals. However, if solutions are not found to mitigate the impact of stellar activity, the detection or confirmation of potential Earth-twins will be extremely challenging and false detections could plague the field.

One of the most challenging stellar signals to characterize and to correct for are the signals induced by stellar activity. Stellar activity is responsible for creating magnetic regions on the surface of stars, and those regions change locally the temperature and the convection, which can induce spurious RVs variations (e.g. Meunier et al. 2010; Dumusque et al. 2014; Borgniet et al. 2015). In theory, it should be easy to differentiate between the pure Doppler-shift induced by a planet, which shifts the entire stellar spectrum, and stellar activity, which modifies the shape of spectral lines and by doing so creates a spurious shift of the stellar spectrum (Saar & Donahue 1997; Hatzes 2002; Kurster et al. 2003; Lindegren & Dravins 2003; Desort et al. 2007; Lagrange et al. 2010; Meunier et al. 2010; Dumusque et al. 2014). However, on quiet GKM dwarfs, the main targets for precise RVs measurements, stellar activity can induce signals of a few

---

m s^{-1}. This corresponds physically to variations smaller than 1/100th of a pixel on the detector, making changes in the shape of the spectral lines challenging to detect. In order to measure such tiny variations, a common approach is to average the information of all the lines in the spectrum by cross correlating the stellar spectrum with a synthetic or an observed stellar template (Baranne et al. 1996; Pepe et al. 2002; Anglada-Escudé & Butler 2012). The result of this operation gives us the cross-correlation function (CCF). To measure the Doppler-shift between different spectra, and therefore to retrieve the RVs of a star as a function of time, the variations of the CCF barycenter are calculated. The barycenter is generally estimated by the mean of a Normal density shape fit to the CCF. Variations in line shape between different spectra, which indicate the presence of signals induced by stellar activity, are measured by analyzing different parameters of models fit to the CCF. Usually the width of the CCF is estimated using the full-width half-maximum (FWHM) of the fitted Normal density and its asymmetry by calculating the CCF bisector and measuring the bisector inverse slope span (BIS SPAN, Queloz et al. 2001).

If a spurious RV signal is induced by activity, generally a strong correlation will be observed between the RV and chromospheric activity indicators like log(R'_{HK}) or H-α (Boisse et al. 2009; Dumusque et al. 2012; Robertson et al. 2014), but also between the RV and the FWHM of the CCF or its BIS SPAN (Queloz et al. 2001; Boisse et al. 2009; Queloz et al. 2009; Dumusque 2016). Therefore a common strategy when fitting a Keplerian signal to a set of estimated RVs in search of a planet is to include linear terms in the model to account for activity, such as the log(R'_{HK}), the FWHM, and the BIS SPAN (Dumusque et al. 2017; Feng et al. 2017). It is also common to add a Gaussian process to the model to account for the correlated noise induced by stellar activity. The hyperparameters of the Gaussian process can be trained on different activity indicators (Haywood et al. 2014; Rajpaul et al. 2015) or directly on the RVs (Faria et al. 2016). It is therefore essential for mitigating stellar activity to obtain activity indicators that are the most correlated with the RVs but also for which we can obtain the best precision.

Several indicators have been developed that can be more sensitive to line asymmetry than the BIS SPAN. In Boisse et al. (2011), the authors developed V_{span}, which is the difference between the RV measured by fitting a Normal density to the upper and the lower parts of the CCF. This CCF asymmetry parameter is shown to be more sensitive than the BIS SPAN at low signal-to-noise ratio (SNR). Figueira et al. (2013) studied the use of new indicators, BIS-, BIS+, bi-Gauss and V_{asy}. The authors were able to show that when using bi-Gauss, the amplitude in asymmetry is 30% larger than when using BIS SPAN, therefore allowing the detection of lower levels of activity. They also demonstrated that V_{asy} seems to be a better indicator of line asymmetry at high SNR, as its correlation with RV is significantly stronger than any other correlation between the previously proposed asymmetry indicators and RV.

In all the methods described above, except bi-Gauss, the RV and the FWHM are derived using a Normal density fitted to the CCF, and the asymmetry is estimated using a separate approach. In this paper we propose to use a Skew Normal (SN) density to estimate with a single fit of the CCF the RV, the FWHM and the asymmetry of the CCF, as this function includes a skewness parameter (Azzalini 1985).

The paper is organized as follows. In Sec. 2 we introduce the SN density, describe its applicability for modelling the CCF and study how the SN parameters relate to the mean of the Normal density, the FWHM, and the BIS SPAN of the CCF. In Sec. 3

we propose a linear model to correct for stellar activity signals in RVs, which extends the linear models previously proposed for this purpose (e.g. Dumusque et al. 2017; Feng et al. 2017). In Sec. 4 the performance of the SN fit to the CCF is investigated using simulations coming from the Spot Oscillation And Planet 2.0 code (SOAP 2.0, Dumusque et al. 2014), followed by an analysis of real observations in Sec. 5. In Sec. 6 error bars are computed for the different estimated CCF parameters. Finally the discussion of the results and the conclusions are presented in Secs. 7 and 8, respectively.

2. The Skew Normal distribution

The Skew Normal (SN) distribution is a class of probability distributions which includes the Normal distribution as a special case (Azzalini 1985). The SN distribution has, in addition to a location and a scale parameter analogous to the Normal distribution's mean and standard deviation, a third parameter which describes the skewness (i.e. the asymmetry) of the distribution. Considering a random variable $Y \in \mathbb{R}$ (where $\mathbb{R}$ is the real line) which follows a SN distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ (i.e., the positive real line), and skewness parameter $\alpha \in \mathbb{R}$, its density at some value $y \in Y$ can be written as

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\frac{\alpha(y - \xi)}{\omega}\right), \qquad (1)$$

where ϕ and Φ are, respectively, the density function and the distribution function of a standard Normal distribution[1]. The skewness parameter α quantifies the asymmetry of the SN. Examples of SN densities under different skewness parameter values and the same location and scale parameters ($\xi = 0$ and $\omega = 1$) are displayed in Fig. 1. A usual Normal distribution is the special case of the SN distribution when the skewness parameter α is equal to zero[2]. For reasons related to the interpretation of the parameters in Eq. 1 and computational issues with estimating α near 0, a different parametrization is used in this work, which is referred to as the *centered parametrization* (CP). This CP is much closer to the parametrization of a Normal distribution, as it uses a mean parameter μ, a variance parameter σ^2 and a skewness parameter γ. In order to define the CP, we need to express the CP parameters (μ, σ^2, γ) as a function of (ξ, ω^2, α). This can be done using the following relations:

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1 - \beta^2), \quad \gamma = \frac{1}{2}(4 - \pi)\beta^3 \left(1 - \beta^2\right)^{-3/2}, \quad (2)$$

where $\beta = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{\sqrt{1 + \alpha^2}}\right)$ (e.g. Arellano & Azzalini 2010).

By using Eq. 2, the new set of parameters (μ, σ^2, γ) provides a clearer interpretation of the behavior of the SN distribution. For the α values used in Fig. 1, the corresponding values of (μ, σ^2, γ) are displayed in Table 1. In particular, μ and σ^2 are the actual mean and variance of the distribution, rather than simply a location and scale parameter, and γ provides a measure of the skewness of the SN. Along with the mean of the SN, we consider the median of the distribution as a measure of its barycenter. See Table 1 for the medians of the SN densities displayed in Fig. 1.

---

[1] A standard Normal distribution is a Normal distribution with a mean of 0 and a standard deviation of 1.

[2] This can be seen from Eq. 1. If $\alpha = 0$ then $\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \Phi(0) = 0.5$ and therefore $SN(y; \xi, \omega, 0) = \frac{1}{\omega} \phi\left(\frac{y-\xi}{\omega}\right)$ which is the density of a Normal distribution. Note that $\Phi(0) = 0.5$ because $\Phi(0)$ is the the probability that a standard Normal random variable is less than or equal than 0.

Fig. 1. Density function of a random variable Y following the SN distribution $SN(\xi, \omega, \alpha)$ with location parameter $\xi = 0$, scale parameter $\omega = 1$ and different values of the skewness parameter α indicated by different colors and line types. Note that the solid black line has an $\alpha = 0$ making it a Normal distribution.

Table 1. CP values (μ, σ^2, γ) along with the median corresponding to the α values shown in Fig. 1, with location parameter $\xi = 0$ and scale parameter $\omega = 1$. Values are rounded to three decimal places.

α	μ	σ^2	γ	Median
-3	-0.757	0.427	-0.667	-0.672
0	0.000	1.000	0.000	0.000
2	0.714	0.491	0.454	0.655
6	0.787	0.381	0.891	0.674
10	0.794	0.370	0.956	0.674

Further details about the parametrization from Eq. 1, called the *Direct Parametrization* (DP), the CP, and general statistical properties of the SN can be found in Azzalini & Capitanio (2014).

2.1. Fitting the Skew Normal density to the CCF

To fit the CCF using a SN density shape, we use a least-squares algorithm and the following model:

$$f_{CCF}(x_i) = C - A \times SN(x_i; \mu, \sigma^2, \gamma), \quad i = 1, \ldots, n \quad (3)$$

where C is an unknown offset for the continuum of the CCF, A is the unknown amplitude of the CCF, sometimes referred to as the CCF contrast, and μ, σ^2 and γ are, respectively, the mean, the variance, and the skewness of the SN as defined above. The values $x_1, \ldots, x_n$ are the different values of the x-axis of the CCF, generally in velocity units (e.g. $m\,s^{-1}$).

When using a Normal density shape model for the CCF, the estimated mean is used as the estimated RV and the FWHM[3] is used to quantify the width of the CCF. Because the Normal density is symmetric, the skewness is not defined and therefore a separate approach is necessary to estimate the skewness of the CCF. An estimated skewness parameter is generally obtained by calculating the BIS SPAN of the CCF (see Sec. 1, and e.g. Queloz et al. 2001).

With the proposed SN approach, we propose two estimators of the RV: the mean and median of the SN model fit (referred to

---

[3] FWHM= $2\sqrt{2\ln 2}\,\sigma$ with standard deviation σ

as SN mean RV and SN median RV, respectively), and present advantages and limitations for both of these choices in Sec. 5 and Sec. 6. The width of the SN, SN FWHM, is defined in the same way as for the Normal density[4], and finally the skewness of the CCF is estimated by the γ parameter.

To evaluate the strength of the correlation between the estimated RVs and the different stellar activity indicators, we calculated the Pearson correlation coefficient, R, which in its general form is defined as:

$$R(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma(x)\sigma(y)}, \quad (4)$$

where x and y are two quantitative variables, $\mathrm{cov}(x, y)$ indicates the covariance between x and y, and $\sigma(x)$ and $\sigma(y)$ represent their standard deviations. A p-value for the statistical test with null hypothesis $H_0 : R = 0$ is also generally provided.

3. Radial Velocity correction for stellar activity

Exoplanets only produce a pure RV signal. On the contrary, stellar activity, and, in particular, the presence of active regions on the stellar photosphere, do not produce blueshifts or redshifts of the entire stellar spectrum but can create spurious RV signals by modifying the shape of spectral lines. To track these variations in the shape of the spectral lines, a common approach consists in using the FWHM, the BIS SPAN, or other indicators such as those introduced in Boisse et al. (2011) or Figueira et al. (2013), which provide information on the width and asymmetry of the CCF. A strong correlation between the estimated RVs and one or more of these parameters provides an indication that stellar activity signals may be affecting the measurements.

When fitting for planetary signals in RV data, it is common to include linear dependencies with the BIS SPAN and the FWHM to take into account the signal induced by stellar activity (e.g. Dumusque et al. 2017; Feng et al. 2017). We propose to add additional parameters in the model to correct for stellar activity: (i) an amplitude parameter A of the CCF (referred to as the CCF contrast) and (ii) an interaction term for γ and SN FWHM (or the BIS SPAN and the FWHM in the Normal case). The stellar activity correction we propose can therefore be written as:

$$RV_{\mathrm{activity}} = \beta_0 + \beta_1 A + \beta_2 \gamma + \beta_3 \mathrm{SN\,FWHM} + \beta_4 (\gamma \mathrm{SN\,FWHM}) + \epsilon, \quad (5)$$

where β_0 is the intercept and ϵ is the random error with mean equal to 0 and covariance matrix equal to $\sigma^2 I$ (I defined as the identity matrix). The contrast parameter A accounts for the presence of a spot on the stellar surface, which produces a change in the amplitude of the CCF, in addition to changes in asymmetry or width (see e.g. Fig. 2 in Dumusque et al. 2014). The benefits of including a variable that quantifies the interaction between γ and SN FWHM (or BIS SPAN and FWHM) will be better understood through the results of the examples presented in Sec. 4. This interaction term can account for possible interactions between SN FWHM (or FWHM) and γ (or BIS SPAN), meaning that each variables' association with the response, RV_{activity}, depends also on the other variable.

The proposed model is analyzed using statistical tests on the parameters β_0, β_1, β_2, β_3 and β_4 where the null hypothesis is $H_0 : \beta_i = 0$, for $i = 0, \ldots, 4$. The significance level for the tests are set at 0.05. The coefficient of determination, R^2, is used

---

[4] Note that SN FWHM does not correspond to the width of the SN density at half maximum like in the Normal case.

to assess how well the proposed linear combination of variables accounts for the variability of RV_{activity}.

The proposed function defined in Eq. 5 is the result of statistical and astronomical considerations. In particular, we considered issues with correlations between the model covariates, which can lead to issues because the matrix needed to calculate the estimates could be singular; this problem is known in statistics as *multicollinearity* and more discussion can be found in Belsley (1991). In the analysis of real data presented in this work, we never observed a correlation coefficient exceeding 0.66 between the asymmetry and width parameters and therefore the problem of multicollinearity appeared to be mitigated. We also investigated the statistical significant of the interactions between A and the width of the CCF, and A and the asymmetry of the CCF; those two interactions were not statistically significant.

4. Simulation Study

In order to evaluate the performance of the proposed SN approach for modelling the CCF and the benefit of using the proposed correction for stellar activity (see Eq. 5), we begin by considering a simulation study using spectra generated from the Spot Oscillation And Planet 2.0 code (SOAP 2.0, Dumusque et al. 2014).

For a given configuration of spots and faculae on the stellar surface, SOAP 2.0 outputs simulated CCFs as a function of rotational phase. The code also returns the RV and the FWHM estimated using a Normal model for the CCF, and the BIS SPAN obtained by calculating the bisector of the CCF. SOAP 2.0 gives noiseless CCFs affected by stellar activity, which are used to compare the performances of a SN and Normal models of the CCF.

For the simulations discussed below, a star similar to the Sun was modelled with a solar disc of one solar radius seen equator-on, and a stellar rotational period set to 25.0 days. The stellar effective temperature is set to 5778 K, and a quadratic limb-darkening relation with linear and quadratic coefficients 0.29 and 0.34, respectively (Oshagh et al. 2013; Claret & Bloemen 2011). In order to make the result of the simulations more comparable to real data obtained with the HARPS spectrograph discussed in Sect. 5, the SOAP 2.0 CCFs were generated with a width of 40 km s^{-1} and considering initial spectra with a spectral resolution of R=115'000.

4.1. Faculae

To see the impact of a facula on the parameters of different models of the CCF, we simulated the effect of an equatorial facula of size 3% relative to the visible stellar hemisphere. The facula is face-on when the phase equals to 0. Note that a 3% faculae is relatively large for the Sun; at maximum activity, large faculae generally have a size of 1% (e.g. Borgniet et al. 2015). In Fig. 2, we compare the barycentric variation of the CCF as measured when fitting a Normal density and using its mean (N mean RV), and when fitting a SN density and taking its mean (SN mean RV) or median (SN median RV). We see that all the different estimates of the CCF barycenter present a signal of similar amplitude, however the signal obtained with SN mean RV is notably different from the two others with a maximum amplitude happening at a different phase.

Correlations between the different RV estimates and the different CCF asymmetry or width estimates are displayed in Fig. 3. The correlation between γ and SN mean RV, and γ and SN median RV are stronger than the correlation between BIS SPAN
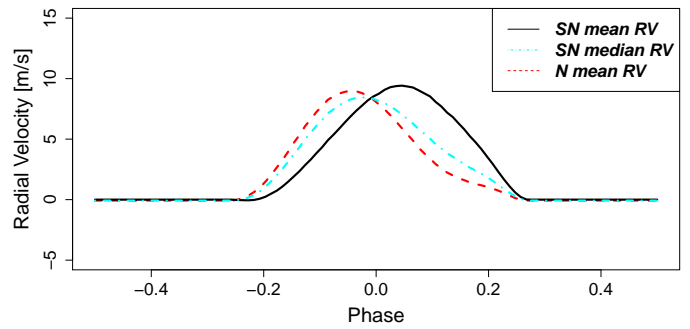
Fig. 2. RV estimates for N mean RV (red dashed line), SN mean RV (black line), and SN median RV (cyan dotted-dashed line). In this case, the CCFs were generated using SOAP 2.0 with an equatorial 3% facula on the simulated Sun. The star does one full rotation between phase -0.5 and 0.5, with the facula being seen face-on for phase 0. The variations observed in SN mean RV are notably different from the variations measured in SN median RV and N mean RV.

and RV, with Pearson correlation coefficient values of R=0.46, -0.67 and -0.09, respectively. There is a stronger correlation between SN FWHM and SN mean RV ($R = 0.98$) than between FWHM and N mean RV ($R = 0.84$), but the correlation between SN FWHM and SN median RV is the weakest ($R = 0.50$). This first analysis shows that in the case of a facula, using some parameters from the SN can lead to stronger correlations than the ones obtained by the usual Normal parameters. Therefore, the SN parameters may better probe stellar activity. We investigate this feature further in the following two sections, where we consider simulated data with a single spot and a spot plus a planet, and in Sec. 5 with real observations.

The RV variations displayed in Fig. 2 are caused only by stellar activity, in this case a facula, so we applied the activity correction proposed in Eq. 5 to evaluate the performance of the model in this setting. The results of this correction are displayed in Fig. 4 and the statistical tests on the coefficients involved in Eq. 5 are summarized in Table 2. The proposed correction for stellar activity is able to account for the majority of the activity signal created by a facula, with a R^2 larger than 0.99. In addition, the rms of the different estimates of the RV reduces from about 3 m s^{-1} before correcting for stellar activity to values below 0.15 m s^{-1} after correcting for stellar activity. We see a slightly smaller rms after correction when using the SN parameters compared to the Normal parameters. When comparing the correction proposed in Eq. 5 with what is generally used (i.e. a linear combination of only the asymmetry and width parameters), we see that the proposed correction is able to reduce the rms of the RV residuals by a factor of 2. Looking at the significance of the coefficients in Table 2, we observe that all the SN or Normal parameters are relevant for the correction.

4.2. Spot

Next we consider the effects on the CCF model parameters due to the presence of an equatorial spot of size 1% relative to the visible stellar hemisphere. The spot is face-on when the phase equals to 0. Note that this is a large spot for the Sun, as large spots are generally on the order of 0.1% (e.g. Borgniet et al. 2015). In Fig. 5, we show the barycentric variations of the CCF induced by this simulated spot. In contrast to the case of the facula, all the different estimates of the CCF barycenter for the spot

Fig. 3. (left) Correlations between the different asymmetry parameters and their corresponding RV estimates in the case of an equatorial 3% facula on the simulated Sun. (right) Correlations between the different width parameters and their corresponding RV estimates for the same facula. In the presence of a facula, both the shape and the width of the CCF change as the star rotates, producing statistically significant correlations.

Table 2. P-values for the estimated coefficients from the model in Eq. 5 for correcting stellar activity induced by an equatorial 3% facula on the simulated Sun. All the parameters corresponding to the Normal or SN variables are statistically significant. The estimated R^2 show that the proposed correction for stellar activity explains the vast majority of the spurious variability present in the different RV estimates.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.033	0.00020	0.61
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.0034	$2.22e-16$	$2.22e-16$
β_3	0.00016	$1.091e-6$	$9.75e-7$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.9978	0.9985	0.9981

have the same shape in variation; however, the amplitude for SN mean RV is slightly smaller.

Fig. 6 shows the correlations between the asymmetry parameters and the different estimates of the CCF barycenter (i.e. SN mean RV, SN median RV and N mean RV). The correlation between γ and SN median RV is the strongest ($R = -0.93$), followed by the BIS SPAN - N mean RV correlation ($R = -0.86$) and the γ - SN mean RV correlation ($R = -0.85$). The correlations between the width and the CCF barycenter are more circular and no substantial correlation is observed. Similar to the case with a facula, some parameters of the SN gives stronger correlations compared to the Normal parameters.

As before, the original RV estimates are corrected using Eq. 5. The results of this correction are displayed in Fig. 7 and the statistical tests on the coefficients involved in Eq. 5 are summarized in Table 3. As with the facula, the proposed correction is able, in the case of the spot, to capture almost entirely the stellar activity signal when considering the SN or Normal parameters, with R^2 values for the linear combination larger than 0.99.

Fig. 4. (top) The spurious estimated RVs (black dots) caused by a facula in the simulated data using a Normal and a SN model, the estimated RVs using Eq. 5 (red crosses), and the estimated RVs using the usual correction for stellar activity (green triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2\text{SN FWHM}$ for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1\text{BIS SPAN} + \beta_2\text{FWHM}$ for the Normal fit. (bottom) The residuals from the model fit using Eq. 5 (red crosses) and the residuals from the usual correction (green triangles). The std. are also reported in the legend, and the residuals have a smaller systematic component when using the proposed model of Eq. 5 compared to the usual model. The tests of statistical significance on the parameters are presented in Table 2.

In Fig. 7, we see that the proposed activity correction is able to reduce the signal of a spot from a raw RV rms larger than $4.80\,\text{m s}^{-1}$ down to a rms smaller than $0.38\,\text{m s}^{-1}$, for any of the different RV estimates. When comparing the activity correction proposed in this paper with what is commonly used, i.e only a linear combination of the width and asymmetry of the CCF, we see that our solution is capable of reducing the RV residual rms by a factor of 3.5, which is even more than the factor 2 found in the case of the facula. The Normal or SN parameters involved in Eq. 5 are all statistically significant to explain the activity signal as seen in Table 3, except the width of the CCF when the RVs are derived with the mean or median of the SN density.

4.3. Spot and planet

The final simulation includes a planetary signal influencing the CCF along with the 1% spot modelled previously (see Sec. 4.2). The purpose of this example is to check if we are able to disentangle these two different sources of variations when using the parameters derived using a Normal or a SN model for the CCF. In this scenario the planet is injected with a semi-amplitude of $10\,\text{m s}^{-1}$ with no eccentricity and with a period corresponding to one-third of the stellar rotational period, i.e. one-third of 25 days.

Fig. 8 shows the variations observed in the CCF barycenter parameters. As in the case of the spot, all RV estimates show similar variations, with SN mean RV showing a slightly smaller amplitude.
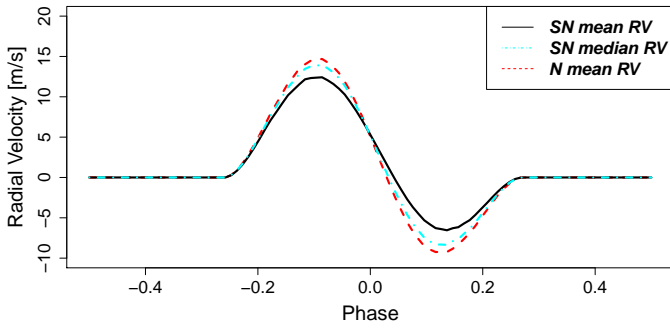
Fig. 5. RV estimates for N mean RV (red dashed line), SN mean RV (black line) or SN median RV (cyan dotted-dashed line) using CCFs generated from SOAP 2.0 with an equatorial 1% spot on the simulated Sun. The star does one full rotation between phase -0.5 and 0.5, with the spot being seen face-on at phase 0. The SN mean RV seems to have the smallest spurious variations caused by the spot.

Table 3. P–values for the different coefficients used in Eq. 5 for the correction of stellar activity induced by an equatorial 1% spot on the simulated Sun. All the parameters corresponding to the Normal or SN parameters are statistically significant to explain the spurious RV variations caused by this spot, except for the width of the CCF when using SN mean RV or SN median RV as RV estimates. The estimated R^2 show that the proposed correction for stellar activity explains the vast majority of the spurious variability seen in the different RV estimates.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.4975	0.21	0.21
β_1	$2e-16$	$2e-16$	$2e-16$
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.017	0.13	0.11
β_4	$2e-16$	$2e-16$	$2e-16$
R^2	0.9959	0.9936	0.9952

The correlation between the different CCF parameters are displayed in Fig. 9. The correlations are weaker than in the case of the spot due to the planet inducing changes in RV without affecting the width or the asymmetry of the CCF. However, the order of the strength of the correlations between the CCF asymmetry parameters and RV are comparable with the ones obtained for the spot-only model: γ–SN median RV has the strongest correlation ($R = -0.84$), followed by the correlation between BIS SPAN–N mean RV ($R = -0.78$) and finally by the correlation between γ–SN mean RV ($R = -0.75$). The patterns observed in the width-RV phase space correlations in Fig. 9 follow a circle, similar to the spot-only model; no substantial correlation is observed between those two parameters.

In order to correct the estimated RVs from the spurious variation caused by the spot, the proposed model for correcting the activity is added to a planetary signal model that takes into account the RV variations caused by a planet. The observed RVs can therefore be modelled as a combination of the activity and the planetary signals:

$$RV = RV_{\text{activity}} + RV_{\text{planet}}, \tag{6}$$

where RV_{activity} can be found in Eq. 5, and RV_{planet}, in the case with no eccentricity, can be modelled by the following sinusoidal function:

$$RV_{\text{exoplanet}} = K \sin\left(\frac{2\pi}{P}(t - t_0)\right), \tag{7}$$

Table 4. P-values for the different coefficients used in Eq. 5 for the correction of stellar activity induced by an equatorial 1% spot on the simulated Sun, and a planet with a period of one-third the rotational period of the star and a semi-amplitude of 10 m s^{-1}. All the parameters corresponding to the Normal or SN variables are statistically significant, except the width of the CCF. Note that since nonlinear least squares was required, the residual standard error rather than the R^2 is displayed as a reference.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.00063	$2e-16$	$1.42e-09$
β_1	$2e-16$	$2e-16$	$2e-16$
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.067	0.40	0.38
β_4	$2e-16$	$2e-16$	$2e-16$
K	$2e-16$	$2e-16$	$2e-16$
P	$2e-16$	$2e-16$	$2e-16$
t_0	$2e-16$	$2e-16$	$2e-16$
Residuals	0.71 m s^{-1}	0.66 m s^{-1}	0.70 m s^{-1}

with amplitude K, orbital period P, and an epoch at the periapsis t_0. The previous three unknown parameters define the planetary orbit.

The proposed model from Eq. 6 was fitted to the RV data and the results of the estimated model are summarized in Table 4. Except for the width parameters with coefficient β_3, all the other Normal or SN parameters are significantly useful to explain the RV variation induced by a spot plus a planet. We also observe that the RV residuals, once corrected for stellar activity and the presence of the planet, are comparable in terms of rms for all the three different RV estimates, with SN mean RV giving a slightly smaller value.

5. Real data application

The simulated SOAP data of the previous section were helpful in assessing the performance of the proposed methodology in a setting where the ground truth was known. In this section we present an analysis conducted on real observations, in particular, the star Alpha Centauri B (Alpha Cen B), and compare the performance when fitting a CCF using the SN density defined in Sec. 2.1 with the commonly employed approach based on fitting a Normal density for estimating the RV and retrieving the BIS SPAN for evaluating the asymmetry of the CCF. Four other stars have been analyzed with the proposed method and details can be found in Appendix A. For all the stars considered in the presented work, only CCFs that were derived from spectra that had at least a SNR of 10 at 550 nm were selected.

5.1. Comparison for Alpha Cen B of the different CCF parameters derived with the Normal and the Skew Normal

A total of 1808 CCFs that were derived from the spectra of Alpha Centauri B taken in 2010 by the HARPS spectrograph have been analyzed. Note that more observations were carried out that year, however only the data that were not significantly affected by contamination from Alpha Centauri A were used (see Dumusque et al. 2012). The selected observations represent arguably, among all RV data existing, the best sampled and most precise RV data set showing strong solar-like activity signal (Thompson et al. 2017; Dumusque et al. 2012).

Fig. 6. (left) Correlations between the different asymmetry parameters and their corresponding RV estimates in the case of an equatorial 1% spot on the simulated Sun. (right) Correlations between the different width parameters and their corresponding RV estimates for the same spot. In the presence of a spot, both the shape and the width of the CCF change as the star rotates. However, only the asymmetry produces a statistically significant correlation with the different RV estimates. The width parameters and their corresponding RV estimates present weak correlations and, in general, much weaker correlations compared to the results obtained when an equatorial 3% facula is present on the simulated Sun.

First, the correlation between γ and BIS SPAN is evaluated. In the left panel of Fig. 10, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 0.00072 and a strong Pearson correlation coefficient of $R = 0.954$. This strong correlation suggests that both γ and BIS SPAN are measuring similar asymmetries for the CCFs. It also provides a conversion between the dimensionless γ parameter into $\mathrm{m\,s^{-1}}$ using the slope of $0.00072 \ \mathrm{m\,s^{-1}}$.

The right plot of Fig. 10 displays the comparison between the RVs estimated using the SN density and the Normal density. The amplitude of the activity signal is slightly stronger for the SN mean RV (in the top-right plot the black circles of the SN mean RV tend to show more variability), while the signal measured using N mean RV or SN median RV are comparable. This behavior is similar to the one observed for the facula simulated with SOAP 2.0 in Sec. 4.1, suggesting that the activity signal could be due to faculae present on Alpha Centauri B. Another

feature that suggests the presence of faculae is the strong positive measured correlation between γ and SN mean RV and between SN FWHM and SN mean RV, as displayed in Fig. 11; as we saw in Sec. 4.2 a spot induces a negative correlation between γ and SN mean RV and weak correlations are measured between SN FWHM and SN mean RV. Dumusque (2014) had previously suggested that the activity of Alpha Centauri B could be due to faculae.

Similar to the analyses presented in Sec. 4, in Fig. 11 we compare the correlations between the asymmetry or the width parameters of the CCF and the RV. For this analysis, we also include the asymmetry parameters derived in Boisse et al. (2011), V_{span} and in Figueira et al. (2013), BIS-, BIS+, Bi Gauss and V_{asy}, as these authors found those asymmetry parameters more correlated with the RVs than BIS SPAN. It is clear in the case of Alpha Centauri B that the correlation found between γ and SN mean RV is the strongest. The Pearson correlation coefficient is
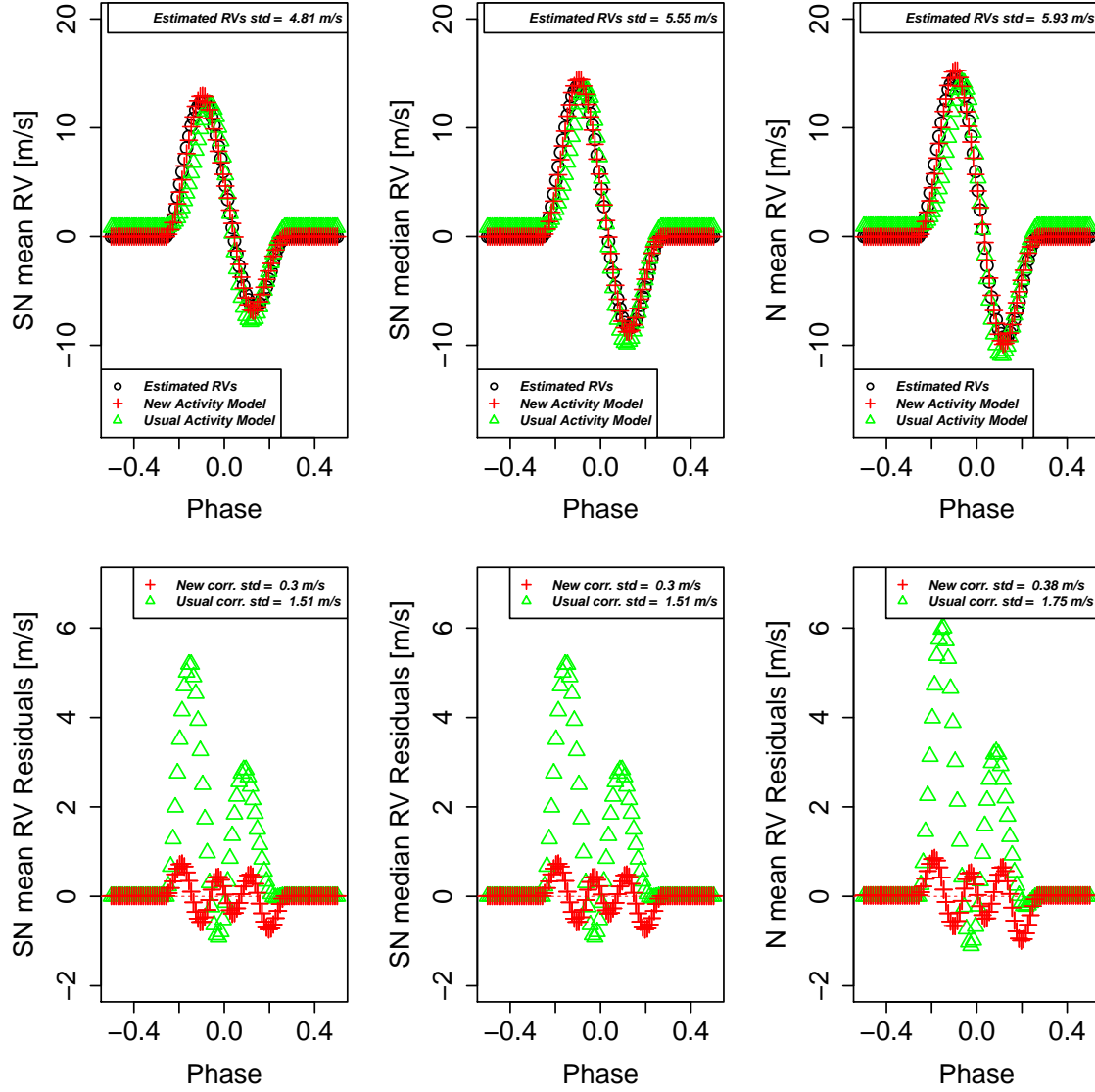
Fig. 7. (top) The spurious estimated RVs (black dots) caused by a spot in the simulated data, the estimated RVs using Eq. 5 (red crosses) and the estimated RVs using the usual correction for stellar activity (green triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1 \gamma + \beta_2 \text{SN FWHM}$ for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1 \text{BIS SPAN} + \beta_2 \text{FWHM}$ for the normal fit. (bottom) The residuals from the model fit using Eq. 5 (red crosses) and the residuals from the usual correction (green triangles). The std. are also reported in the legend, and the residuals have a smaller systematic component when using the proposed model compared to the usual model. The tests of statistical significance on the parameters are presented in Table 3.

$R = 0.74$, while the next strongest is $R = 0.42$ for all the other asymmetry-N mean RV correlations. The correlations between the width and the RV estimates for Alpha Centauri B is also the strongest for the SN parameters, with $R = 0.82$ for SN FWHM-SN mean RV, compared to $R = 0.70$ for FWHM-N mean RV.

Results illustrating the performance of the stellar activity correction proposed in Sec. 3 are displayed in Fig 12. For Alpha Centauri B, the RV estimated with SN mean RV has a std that is 35% larger than the std of the RV estimated with the N mean RV, and the std of SN median RV is 9% larger than that of the N mean RV. Even though we see these differences in the estimated RV, once we correct for stellar activity using Eq. 5, the rms of the residuals are essentially the same for all three approaches. Although the correlations between the different parameters from the SN density are more sensitive to stellar activity than those obtained with a Normal density fit, the proposed linear model

that corrects for stellar activity does not necessarily perform better in the SN case than in the Normal case. The new correction for stellar activity proposed in Sec. 3 performed similarly to the usual correction that uses only a linear combination of the width and the asymmetry of the CCF, however it still gives RV rms that are 8 cm s^{-1} smaller than the usual correction.

The results of the statistical tests of the different parameters used for correcting activity can be found in Table 5. The BIS SPAN (coefficient β_2) is not statistically significant for the parameters derived from the Normal density fit. However, all the other parameters in the Normal and SN cases are statistically significant for modelling stellar activity. By analyzing the values of the coefficient of determination, R^2, we see that the model for the SN mean RV is able to capture the highest percentage of variability in the estimated RV. This is not a surprising result since the three different RV estimates have the same RV residual rms
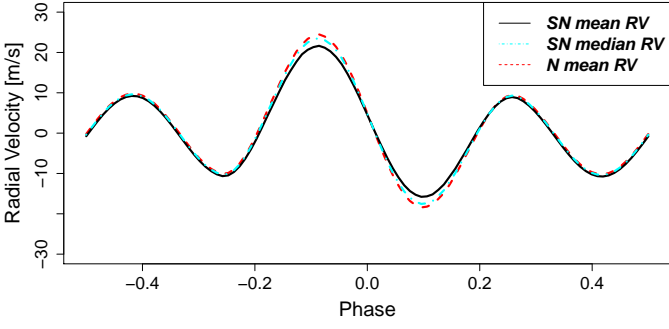
Fig. 8. RV estimates for N mean RV (red dashed line), SN mean RV (black line) or SN median RV (cyan dotted-dashed line). In this case, the CCFs have been generated using SOAP 2.0, considering an equatorial 1% spot on the simulated Sun in addition to a planet with a period of one-third of the rotational period of the star and with an amplitude of $10 \, \text{m s}^{-1}$. The star does one full rotation between phase -0.5 and 0.5, with the spot being seen face-on at phase 0.

Table 5. P-values for the different coefficients used in Eq. 5 for the correction from stellar activity in Alpha Centauri B data. All the variables corresponding to the Normal or SN parameters are statistically significant, except for the asymmetry of the CCF when using the BIS SPAN with the N mean RV. The R^2 shows that the proposed linear combination explains the most variability in RVs due to the stellar activity when the RVs are estimated with the SN mean RV.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.49	0.90	0.027
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.33	$2.22e-16$	$1.23e-11$
β_3	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.57	0.78	0.66

after correction for activity, but before correction, the SN mean RV had the largest RV rms (see Fig 12).

5.2. Comparison for HD192310, HD10700, HD215152 and Corot-7 of the different CCF parameters derived with the Normal and the Skew Normal

In the previous section we evaluated the improvement obtained by the SN parameters compared to the Normal parameters and the BIS SPAN. We carryout similar analyses for four other main sequence stars: HD192310 (K2V, Pepe et al. 2011), HD10700 (G8V, Feng et al. 2017), HD215152 (K3V, Delisle et al. 2018) and finally Corot-7 (K0V, Haywood et al. 2014). The same correlations plots and residual plots displayed in the previous section for Alpha Centauri B can be found found for those new four stars in Appendix A.

The correlations between the parameters of these additional stars are similar to those obtained for Alpha Centauri B. The correlation between γ and SN mean RV is the strongest among all the asymmetry-RV correlations. Between the width parameters and the estimated RV, the strongest correlation often is between SN FWHM and SN mean RV. However, there is one exception in the case of HD10700 where the Pearson correlation coefficient between FWHM and N mean RV is equal to $R = 0.53$, while it is $R = 0.42$ between SN FWHM and SN mean RV, and $R = 0.5$ between SN FWHM and SN median RV.

Except for the special case discussed above for HD10700, the analyses of those four stars, in addition to the analyses on Alpha Centauri B, show that the parameters derived when using a SN density are generally more sensitive to activity. Therefore using the SN parameters, and in particular estimating RV using SN mean RV, can carryout to better detection of stellar activity over the Normal parameters. More specifically, this is the case for the evaluation of the asymmetry-RV correlations for Alpha Centauri B, HD10700, HD215152, HD192310 and Corot-7, and the width-RV correlation for Alpha Centauri B, HD215152, HD192310 and Corot-7 (see Appendix A).

When correcting for stellar activity for Alpha Centauri B, although the uncorrected RV rms was larger for SN mean RV (compared to the RVs obtained using N mean RV), once corrected for activity using the new model proposed in Sec. 3, both RVs estimates had similar residuals. For HD10700, HD215152, and HD192310, the proposed and usual models were giving similar RV residual rms. However, for Corot-7, the new correction is able to provide RV residual rms 23 cm s^{-1} smaller than the one obtained with the usual correction.

5.3. Detection limits when using the estimated RVs from the Normal or the Skew Normal models

In the previous section, we saw that the estimated RV measured when considering a SN or a Normal density resulted in different amplitudes, especially when using the SN mean RV. However, once corrected for stellar activity using the linear combination presented in Eq. 5, as shown in the bottom plots of Fig 12, the rms of the residuals are essentially the same for all three approaches. In this section, we investigate the ability of the three different RV estimators (N mean RV, SN mean RV, and SN median RV) to detect planetary signals among stellar activity, and also compare them when using the usual stellar activity correction with the proposed stellar activity model of Eq. (5). To carryout this test, the minimum detected amplitude of an injected planetary signal is estimated at different orbital periods when considering data affected by stellar activity.

In order to obtain CCFs affected by realistic stellar activity signals, the CCFs from Alpha Centauri B used previously were considered. To simulate a planetary signal, the CCFs were blue- or red-shifted with the desired amplitude, period, and phase. Several RV data sets with the same stellar signal, but different planetary signals were generated using parameters corresponding to the following grid:

- period of 3, 5, 7, 9, 11, 15, 20, 25, and 30 days,
- amplitude from 0.5 to 3 m s^{-1} by steps of 0.05 m s^{-1},
- 10 different phases, evenly sampled between 0 and 2π.

For each of the 4500 simulations we derived at first the corresponding three set of RV, namely N mean RV, SN mean RV and SN median RV. [[**Xavier:** On each of those RV sets, we performed an analysis similar to Sec. 4.3, i.e. fitting the activity signal using Eq. (5) or the usual correction along with a circular planetary signal (see Eq. (6)). Because of the non-linearity of the model to account for the planet, we used a non-linear least squares for the fit (Levenberg 1944; Marquardt 1963; Teunissen 1990). Such a model requires initial conditions close to the real solution, otherwise the algorithm can converge to a local minimum. Because our goal here is to compare the planetary detection limits using the three different RV estimates and the two different activity model proposed, and not to discuss what is the best method to explore the parameter space, we initialised the minimisation algorithm to the real period of the planetary signal

Fig. 9. Evaluation of the correlation between the RVs and the asymmetry parameters of the simulated data with a 1% spot and an injected planetary signal. The shape of the CCF changes as the spot moves, producing statistically significant correlations only between the estimated RVs and the asymmetry parameter. The correlations between the estimated RVs and the width parameter of the CCF are weaker than the case with only a spot.

injected to avoid being blocked in a local minimum. We also selected as initial amplitude the peak-to-peak amplitude of the RV data set. The argument of periapsis t_0 was initialised to the time when the RV were crossing 0 since we use a sinusoidal function to fit the planetary signal (see Eq. (7)).]]

Once the parameters involved in Eq. (6) have been estimated, signals in the residuals, defined as $RV - RV_{activity}$, were analyzed using a Generalized Lomb-Scargle periodogram (Lomb 1976; Scargle 1982; Zechmeister & Kürster 2009). If a signal with a P-value[5] smaller than 1% had a period compatible with the injected planetary period within an error budget of 20%, the signal was considered significant and the corresponding planet considered detected. [[**Xavier:** For each period considered, we searched for the minimum amplitude at which at least 80% of the planets with different phases were detected. This minimum amplitude detected as a function of period is shown in Fig. 13 for the three different RV estimates (N mean RV, SN mean RV, and

SN median RV) when using the new stellar activity correction proposed in this paper, and when using the usual activity correction.]] We can see that the new correction for stellar activity based on Eq. (5) slightly improves the detection limit of the exoplanet compared to the usual approach, and the three estimators of RV give similar detection limits. These results suggest that any of the RV estimators can be used when searching for a planetary signal in RV data contaminated by stellar activity, though additional investigation is needed.

Although the detection limits remain relatively constant for the periods ≤20 days, there is a notable increase in the detection limit at 25 days. This is likely caused by the fact that the periods of the simulated planets at 25 days are close to the first harmonic or to the rotational period of the star (36.7 days, Dumusque et al. 2012) and therefore close to the semi-periodicity of the stellar activity signal. [[**Jessi:** wouldn't a period of 30 be considered closer to the rotation period and so should also be higher based on this reasoning?]]

---

[5] The P-values were estimated using a bootstrap procedure.

Fig. 10. (left) Correlation between γ and the BIS SPAN for Alpha Centauri B. The strong correlation suggests these two parameters are similarly measuring the asymmetry. (top right) RVs as function of Julian Day for Alpha Centauri B in 2010. The RVs are estimated using the mean of a Normal fitted to the CCF (red triangles), or the mean (black circles) or median (cyan pluses) of a SN density fitted to the CCF. (bottom right) Differences between the RVs estimated with the Normal density and those from the SN density.

6. Estimation of standard errors for the CCF parameters

In this section, we investigate how the photon noise influences the CCF parameters derived either by a Normal density or a SN density fit. Because a CCF is obtained from a cross-correlation, each point of a CCF is correlated with the other points. Therefore, we cannot simply vary each point in the CCF by their respective error bars and then recalculate the best SN or Normal density fit to see how the CCF noise influences the estimation of the parameters of interest (i.e., N mean RV, SN mean RV, SN median RV, FWHM, SN FWHM, BIS SPAN and γ). Instead, we go to the individual spectrum where each individual points can be considered independent from the others. The standard error on each point of a spectrum is given by the photon noise, which follows a Poisson distribution and is therefore estimated by taking the square-root of the measured flux.

The following method was carried out in order to estimate the error bars on the different parameters derived from the CCF. We first modify the values of all the points in the spectrum given their respective error bars. To do so random Gaussian noise with standard deviation the square-root of the flux was added across each spectrum. The CCF was calculated using this spectrum according to the method presented in Pepe et al. (2002), then fit by either a Normal or SN density with the parameters recorded. This process was repeated a hundred times in order to obtain a distribution for each CCF parameter, and the standard deviations of the resulting distributions provide estimates of the standard errors for the CCF parameters.

The standard errors were computed for each CCF parameter for the HARPS measurements of HD215152, HD192310 and Corot-7. These three stars include measurements that cover the range of SNR measured at 550 nm (SNR550) from 10 to 500,

which represent the very low SNR limit and the saturation limit of the HARPS detector, respectively. HD10700 and Alpha Centauri B were not included because they have a large number of measurements, which would require a substantial computational effort. The variation of the noise for each CCF parameter as a function of SNR550 is displayed in Fig. 14. The top row shows the standard errors of the three different estimated RV's, the width, and the asymmetry estimates. Note that because BIS SPAN and γ do not have the same units, the estimated slopes of the correlation between those two parameters to transform γ in $m\,s^{-1}$ were used (see Fig. 10 and Table A.1 for the value of the slope for each star). The bottom row shows the ratio between the standard errors measured when using the SN parameters and the Normal parameters. Values smaller (larger) than one will imply that standard errors from the SN parameters are more (less) precise than the Normal parameters.

The different RV estimates all appear to follow a similar exponential decay. Considering each RV estimate individually, it appears that all the measurements follow the same curve even though the measurements are from three different stars. This suggests that for the stars considered, the precision in each RV estimate may be driven by the SNR of the analyzed spectra. This is not surprising as the three stars studied have very similar spectral types; they are all main sequence K-dwarfs.

When comparing the three different estimates for the RV, the SN mean RV has standard errors that are 60% larger than the N mean RV. However, the SN median RV gives errors 10% more precise than N mean RV. The parameters describing the width of the CCF, FWHM and SN FWHM, have comparable standard errors. Finally, for the asymmetry parameters, γ has standard errors that are 15% more precise than BIS SPAN. In conclusion, when fitting a SN density to the CCF and using SN median RV as the RV estimate, we are able to improve the precision on the
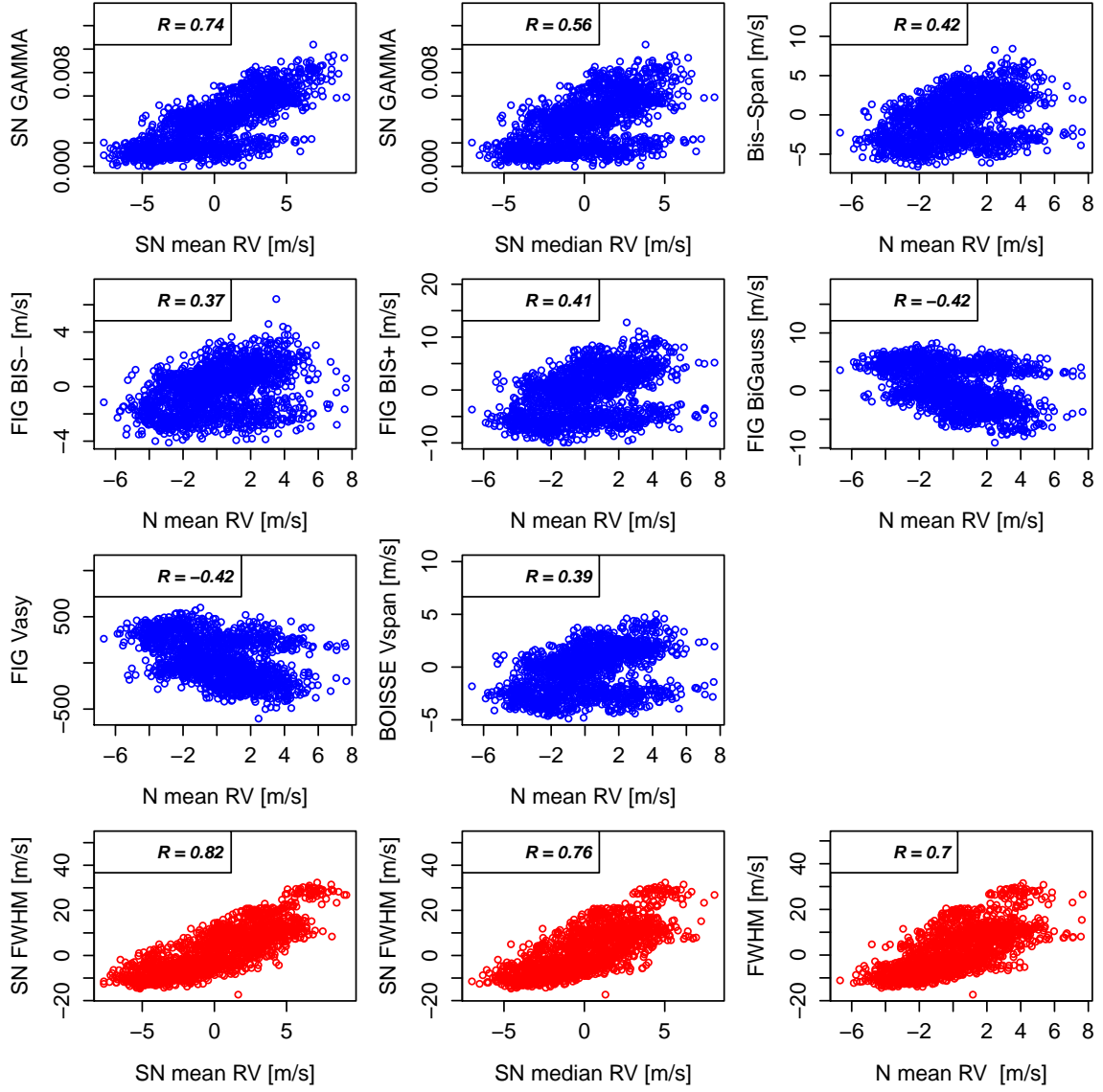
Fig. 11. (top three rows) Correlations between the asymmetry parameters and their corresponding estimated RVs for Alpha Centauri B. (bottom row) The correlation between the FWHM and the estimated RVs. The correlations are stronger when using parameters derived from the SN fit than the Normal one. The estimated *R*'s are all statistically significant.

estimated RV by 10%. Using the SN density, we are also able to improve by 15% the precision on the estimated asymmetry parameter of the CCF. However, the SN mean RV should not be use to derive precise RV estimates as the precision on this parameter is 60% worse than the precision on the RVs derived from N mean RV.

7. Discussion

When fitting a SN density shape to the CCF, parameters used to define the RV (i.e. the CCF barycenter), the amplitude (sometimes called the CCF contrast), the width and the asymmetry of the CCF can all be estimated in a single model framework. For the estimation of the RV, we investigated the use of the mean and the median of the SN density. The width is derived using the variance of the SN density (SN FWHM= $2\sqrt{2ln(2)\sigma^2}$) and the asymmetry by using γ, skewness parameter of the SN density.

To evaluate the performance of the proposed SN framework, tests on both simulated and real data were carried out and compared to the commonly employed approach of fitting a Normal density shape (for estimating the RV and FWHM) and then separately deriving the BIS SPAN (for estimating the asymmetry). The simulated CCFs were generated using the SOAP 2.0 code, which can simulate activity signals induced by a spot or a facula on a solar-like star. The results of the simulation study suggest that the parameters derived from the SN density fit are more sensitive to activity than the parameters obtained by the usual Normal method, making them more useful indicators of activity. In this case, sensitivity was measured using the correlation between the asymmetry parameter and the estimated RVs; both SN mean RV and SN median RV had a stronger correlation with γ than the correlation between N mean RV and BIS SPAN. Moreover, the correlation between the FWHM and the estimated RVs is also stronger when using the parameters from the SN compared to the parameters from the Normal. The SN parameters contin-

Fig. 12. (top) The RVs (black dots) for Alpha Centauri B estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std, black dots) and the residuals from the usual correction (Usual corr. std, blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1 \gamma + \beta_2 \text{SN FWHM}$ for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1 \text{BIS SPAN} + \beta_2 \text{FWHM}$ for the normal fit. The residuals have a smaller systematic component when using the proposed model of Eq. 5 (black dots) compared to the usual model (blue triangles).

ued to have stronger correlations than the Normal in the setting where a planetary signal was added the SOAP 2.0 with a single spot.

The conclusion found for the SOAP 2.0 simulations that the correlation between the CCF asymmetry and the estimated RV, as well as between the CCF width and the estimated RV, is stronger for the SN parameters than the Normal ones also holds when using real data. However, the strongest asymmetry-RV correlation for the simulated data is between γ and SN median RV, while it is between γ and SN mean RV for the real data. There could be several reasons for this difference. First, noting that SOAP 2.0 is a simulation and, hence, is an imperfect representation of reality so it is not unexpected that there are differences between the simulated and real data. For example, some differences may be due to the fact that SOAP 2.0 uses the spectrum of a spot as input to model the activity induced by a facula. Because the temperature between a spot and a facula is significantly dif-

ferent, their spectra should be different. Additionally, there are expected to be multiple active regions on a star at different locations in longitude and latitude, while the SOAP 2.0 data used in the simulation study included only a single active region on the equator. Nevertheless, in both the simulated and the real cases, the parameters derived from the SN are generally more sensitive to activity than the parameters derived from the Normal. Also, the correlation between the asymmetry and the SN mean RV is consistently stronger than the parametrization of the CCF presented in Boisse et al. (2011) and Figueira et al. (2013). Because the apparent RV signal induced by activity results in a stronger correlation with the SN parameters than between the apparent RV signal and the FWHM of the CCF or its BIS SPAN, this suggests the SN model of the CCF can lead to a better understanding of the spurious variations in RV caused by stellar activity.

Considering the different RV estimates of the real data, the amplitude of stellar activity tends to be largest for the SN mean
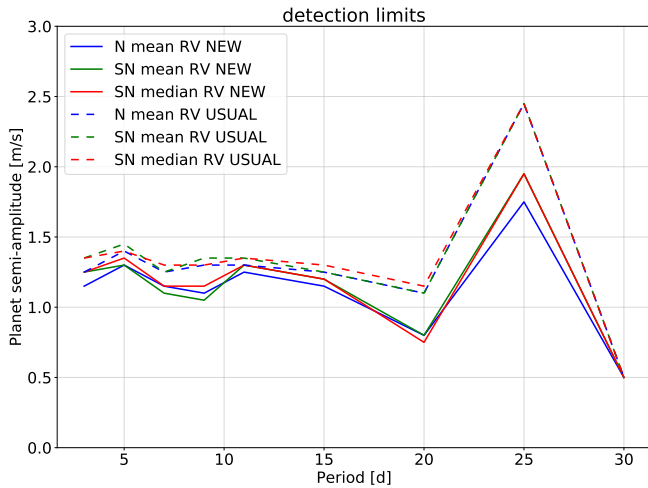
Fig. 13. Detection limits of planetary signals once the stellar activity signal is removed from the raw RVs using the model proposed in Eq. 5 (solid lines) or the usual correction based on $RV_{activity} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{activity} = \beta_0 + \beta_1$BIS SPAN + β_2FWHM for the normal fit (dashed lines). The correction for stellar activity based on Eq. 5 slightly improves the detection limit and the different RV estimators have similar detection limits. [[**Jessi:** to be consistent with other figures, "NEW" and "USUAL" should only have the first letter uppercase.]]

RV, followed by the SN median RV and N mean RV that are extremely similar. This implies that the mean of the SN density appears to be more sensitive to variation in the CCF shape than the median of the SN or the mean of the Normal.

Having an estimator of RV that is more sensitive to stellar activity, such as the SN mean RV, can also help to better probe stellar rotational periods or to better understand the covariance of stellar signals when fitting a Gaussian Process to the RVs (e.g. Faria et al. 2016; Haywood et al. 2014). We saw that the SN mean RV estimator is 60% noisier than the N mean RV estimator. This is not necessarily a negative aspect. For example, in the case where photon-noise is not dominant, stellar activity can be better characterized if its effect can be amplified by measuring the SN mean RV, even if the uncertainties related to this parameter are increased by 60%. [[**Jessi:** I'm not sure I agree with the previous statements - more uncertainty means you would be less certain if the RV estimate is due to actual stellar activity or just photon noise. It is also stated throughout that precise estimates are needed (e.g. in the conclusion)]] [[**Umberto:** Xavier, can you please develop a little bit this statement? If I understood correctly, since the correlation between the width and the asymmetry of the CCF with SN mean RV is almost always the strongest, if the star is particularly active we are willing to "sacrifice" the precision of the RV if we can end up with a clear indication of strong activity by the star. Is it correct?]]

We also propose a new model to correct the estimated RV data for stellar activity signals. Generally, when fitting for planetary signals, it is common to use a model composed of one or several Keplerian signals to account for the planets, in addition to a linear combination of the FWHM and BIS SPAN to account for stellar activity signals. The proposed model adds to the linear model a term to account for the amplitude of the CCF and an interaction term between the estimated asymmetry and the width parameters. Using the simulated data from SOAP 2.0, this new model reduces the effect of the stellar activity signal by factors

of about 2 and 3.5 over the usual model for the facula and spot, respectively.

Even if the different RV estimators result in different amplitudes, once the proposed correction for stellar activity is applied, the residuals of the model have similar rms. When comparing the activity correction proposed in this paper with the usual correction that only uses a linear combination of the CCF asymmetry and width, for the simulations based on the presence of a facula or a spot the new proposed correction almost entirely explains the spurious variations in RV. However, when moving to real data, there is just a slight improvement by using the proposed correction function for stellar activity. A further analysis can be performed to see if certain components of the model are not relevant and that can therefore be removed.

A test was carried out to see if some RV estimates were better at finding planets in RV data affected by observed stellar signals. The new correction based on Eq. 5 proposed in this paper to mitigate the effect from stellar activity slightly improves the detection limit respect to the usual one based on $RV_{activity} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{activity} = \beta_0 + \beta_1$BIS SPAN + β_2FWHM for the normal fit. Concerning the definition of the RV using the SN or the Normal fit, all three of the different RV estimators give similar detection limits. Therefore it seems that any of the RV estimate can be used to search for planetary signals.

Finally, we investigated the precision of each of the SN and Normal parameters including the BIS SPAN. It turns out that the SN mean RV should not be used to get precise RVs as the standard errors on this parameter is 60% greater than for N mean RV. However, the SN median RV is 10% more precise than the N mean RV. Regarding the asymmetry estimates, we observe that γ has a precision 15% better than the BIS SPAN.

8. Conclusion

When searching for low-mass exoplanets using the RV technique, it is necessary to retrieve precise estimates of the RV and also to account for variations induced by stellar activity in order to avoid false detections. Stellar activity such as spots and faculae can lead to shape variations in the spectra features, which then results in shape variations of the CCF. The correlations between the width or asymmetry of the CCF and the estimated RV are commonly used as a way to detect if the RVs are affected by stellar activity signals. Because the presence of real planets would result in only a shift in the CCF (not a change of its shape), strong correlations between the shape features of the CCF and the estimated RVs suggest that stellar activity may be present.

In this paper, a new approach for quantifying shape changes in the CCF is proposed using the SN density, which can be used to estimate with a single fit the RV, the width and the skewness of the CCF. This new method is compared to a commonly used method based on a Normal density fit to the CCF. The mean of the Normal density is used as the estimated RV and the FWHM estimates the width of the CCF. Because the Normal density does not have any skewness, another method is necessary to estimate the asymmetry of the CCF, such as the often employed BIS SPAN. In addition, the proposed SN approach is compared to other parameterizations of the CCF asymmetry that have been shown to be sensitive to activity signals (Boisse et al. 2011; Figueira et al. 2013).

In the different tests carried out for this work, the SN parameters performed at least as well as, and most of the times better than the parameters from the Normal approach and the
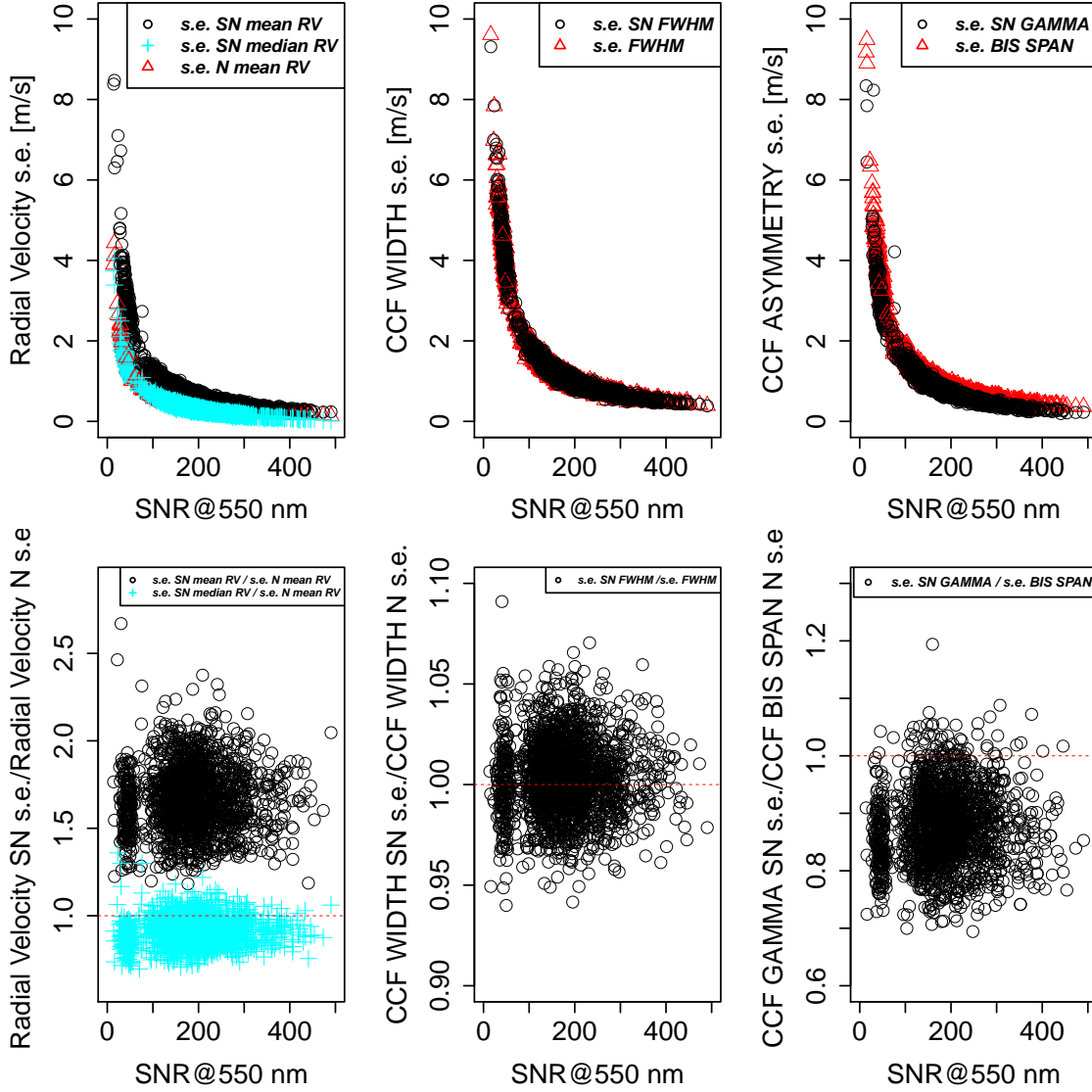
Fig. 14. *Bootstrap analyses on the stars HD215152, HD192310 and Corot-7.* (top) Comparison between the standard errors from the bootstrap analysis of the estimated RVs, FWHM, and asymmetry parameters using the SN fit and the common strategy (Normal fit and BIS SPAN). (bottom) Ratio between the standard errors retrieved on the parameters derived from the common strategy and the corresponding standard errors retrieved on the parameters derived from the SN fit. When using SN mean RV (black circles), the standard errors are in average 60% larger than the standard errors of N mean RV (red triangles). However, the standard errors for SN median RV (cyan crosses) are on average 10% smaller than the standard errors coming from the N mean RV. The use of the asymmetry SN parameter γ leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in m s^{-1}. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

BIS SPAN. The SN parameters γ, SN FWHM, and SN mean RV consistently had stronger correlation than those between any of the parameters derived by the Normal and the BIS SPAN, or the different asymmetry parametrizations presented in Boisse et al. (2011) and Figueira et al. (2013). This suggests the SN parameters may be better at probing stellar activity signals than the other methods. In addition, the uncertainties measured on SN median RV and γ are respectively 10% and 15% smaller than the corresponding uncertainties on N mean RV and BIS SPAN, though the SN mean RV had uncertainties 60% greater than the N mean RV.

Because of the advantages of using the proposed SN approach over the commonly employed approach based on the Normal density fit to the CCF and the BIS SPAN or the asym-

metry parameters described in Boisse et al. (2011) and Figueira et al. (2013), a SN density model for the CCF may be more useful for detecting stellar activity than the previously proposed parametrizations. Correlations between γ and the SN mean RV, and between the width and the SN mean RV can be used to probe stellar activity signals in RV data, and the SN median RV can be used to estimate RV. [[**Umberto:** We also proposed a new model to correct the estimated RV data for stellar activity signals, by using the amplitude of the CCF and an interaction term between the estimated asymmetry and the width parameters. Using simulated data from SOAP 2.0, this new proposed correction reduces the effect of the stellar activity signal by factors of about 2 and 3.5 over the usual model respectively for the facula and spot. When using real data and when injecting a fake planets, the de-

tection limit slightly improved if using the new correction for stellar activity over the usual one.]] We therefore strongly encourage the community to use this proposed SN parametrisation of the CCF.

9. Acknowledgements

Appendix A: Appendix

In this Appendix, a similar analysis as the one presented in Sec. 5 is discussed for four main-sequence stars: HD192310 (K2V, Pepe et al. 2011), HD10700 (G8V, Feng et al. 2017), HD215152 (K3V, Delisle et al. 2018) and finally Corot-7 (K0V, Haywood et al. 2014). The latest HARPS data for these stars can be found on the ESO archive.

Table A.1 summarizes the results obtained by the SN and Normal density models of the CCF. These results are consistent with those from the analysis of Alpha Centauri B. The correlation between γ and SN mean RV is stronger than the correlation between the BIS SPAN and N mean RV or between the asymmetry parameters derived in Boisse et al. (2009) and Figueira et al. (2013) and N mean RV for all the considered stars. The correlation between SN FWHM and SN mean RV is stronger than the correlation between FWHM and N mean RV for three of the four stars. Also for all these stars, the originally estimated RVs were corrected from spurious variations caused by stellar activity using Eq. 5 and Fig. A.2, A.4, A.6, and A.8, show the corrected RVs. Once corrected from stellar activity, the Normal and SN residuals are comparable for the stars HD192310, HD10700 and HD215152. However, the rms of the residuals for Corot-7 are $0.23 \mathrm{~m\,s}^{-1}$ lower for the SN model than the Normal model. The average SNR at 550 nm for the stars HD10700, HD192310, HD215152 and Corot-7 are respectively 273, 207, 141 and 44. Corot-7 has therefore on average a much lower SNR at 550 nm than the others stars, which could be a potential explanation for this small improvement. However further tests on several stars should be perform to confirm this statement.

References

Anglada-Escudé, G. & Butler, R. P. 2012, The Astrophysical Journal Supplement Series, 200, 15

Arellano, R. B. & Azzalini, A. 2010

Azzalini, A. 1985, Scandinavian journal of statistics, 171

Azzalini, A. & Capitanio, A. 2014, The skew-normal and related families. Institute of Mathematical Statistics Monographs

Baranne, A., Queloz, D., Mayor, M., et al. 1996, Astronomy and Astrophysics Supplement Series, 119, 373

Belsley, D. A. 1991, Conditioning diagnostics: Collinearity and weak data in regression No. 519.536 B452 (Wiley New York)

Boisse, I., Bouchy, F., Hébrard, G., et al. 2011, Astronomy & Astrophysics, 528, A4

Boisse, I., Moutou, C., Vidal-Madjar, A., et al. 2009, Astronomy & Astrophysics, 495, 959

Borgniet, S., Meunier, N., & Lagrange, A.-M. 2015, A&A, 581, A133

Claret, A. & Bloemen, S. 2011, A&A, 529, A75

Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in Proc. SPIE, Vol. 8446, 84461V

Delisle, J.-B., Ségransan, D., Dumusque, X., et al. 2018, A&A, 614, A133

Desort, M., Lagrange, A.-M., Galland, F., Udry, S., & Mayor, M. 2007, Astronomy & Astrophysics, 473, 983

Dumusque, X. 2014, ApJ, 796, 133

Dumusque, X. 2016, Astronomy & Astrophysics, 593, A5

Dumusque, X., Boisse, I., & Santos, N. 2014, The Astrophysical Journal, 796, 132

Dumusque, X., Borsa, F., Damasso, M., et al. 2017, Astronomy & Astrophysics, 598, A133

Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Nature, 491, 207

Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. 2011, Astronomy & Astrophysics, 525, A140

Faria, J., Haywood, R., Brewer, B., et al. 2016, Astronomy & Astrophysics, 588, A31

Feng, F., Tuomi, M., & Jones, H. R. 2017, Astronomy & Astrophysics, 605, A103

Feng, F., Tuomi, M., Jones, H. R. A., et al. 2017, AJ, 154, 135

Figueira, P., Santos, N., Pepe, F., Lovis, C., & Nardetto, N. 2013, Astronomy & Astrophysics, 557, A93

Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, Publications of the Astronomical Society of the Pacific, 128, 066001

Hatzes, A. P. 2002, Astronomische Nachrichten, 323, 392

Haywood, R., Collier Cameron, A., Queloz, D., et al. 2014, Monthly notices of the royal astronomical society, 443, 2517

Kurster, M., Endl, M., Rouesnel, F., et al. 2003, ASTRONOMY AND ASTROPHYSICS-BERLIN-, 403, 1077

Lagrange, A.-M., Desort, M., & Meunier, N. 2010, Astronomy & Astrophysics, 512, A38

Levenberg, K. 1944, Quarterly of applied mathematics, 2, 164

Lindegren, L. & Dravins, D. 2003, Astronomy & Astrophysics, 401, 1185

Lomb, N. R. 1976, Astrophysics and Space Science, 39, 447

Marquardt, D. W. 1963, Journal of the society for Industrial and Applied Mathematics, 11, 431

Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20

Meunier, N., Desort, M., & Lagrange, A.-M. 2010, Astronomy & Astrophysics, 512, A39

Oshagh, M., Boisse, I., Boué, G., et al. 2013, A&A, 549, A35

Pepe, F., Lovis, C., Ségransan, D., et al. 2011, A&A, 534, A58

Pepe, F., Mayor, M., Galland, F., et al. 2002, Astronomy & Astrophysics, 388, 632

Pepe, F., Molaro, P., Cristiani, S., et al. 2014, Astronomische Nachrichten, 335, 8

Queloz, D., Bouchy, F., Moutou, C., et al. 2009, Astronomy & Astrophysics, 506, 303

Queloz, D., Henry, G., Sivan, J., et al. 2001, Astronomy & Astrophysics, 379, 279

Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2269

Robertson, P., Mahadevan, S., Endl, M., & Roy, A. 2014, Science, 1253253

Saar, S. H. & Donahue, R. A. 1997, The Astrophysical Journal, 485, 319

Scargle, J. D. 1982, ApJ, 263, 835

Teunissen, P. 1990

Thompson, A., Watson, C., de Mooij, E., & Jess, D. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 468, L16

Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577

Table A.1. Notable correlations between the asymmetry or the FWHM parameters and the RVs for four stars: HD192310, HD10700, HD215152 and Corot 7. The complete results of the analyses of the correlations for the four stars are presented in Fig. A.1, A.3, A.5, and A.7.

Star	# CCFs	R(SN γ, Bis-Span)	slope(SN γ, Bis-Span)	R(SN γ, SN mean RV)
HD192310	1577	0.888	0.000786	0.669(0.64; 0.695)
HD10700	7928	0.78	0.000604	0.322(0.302; 0.342)
HD215152	273	0.763	0.000794	0.571(0.485; 0.646)
Corot 7	173	0.814	0.000607	0.561(0.450; 0.656)

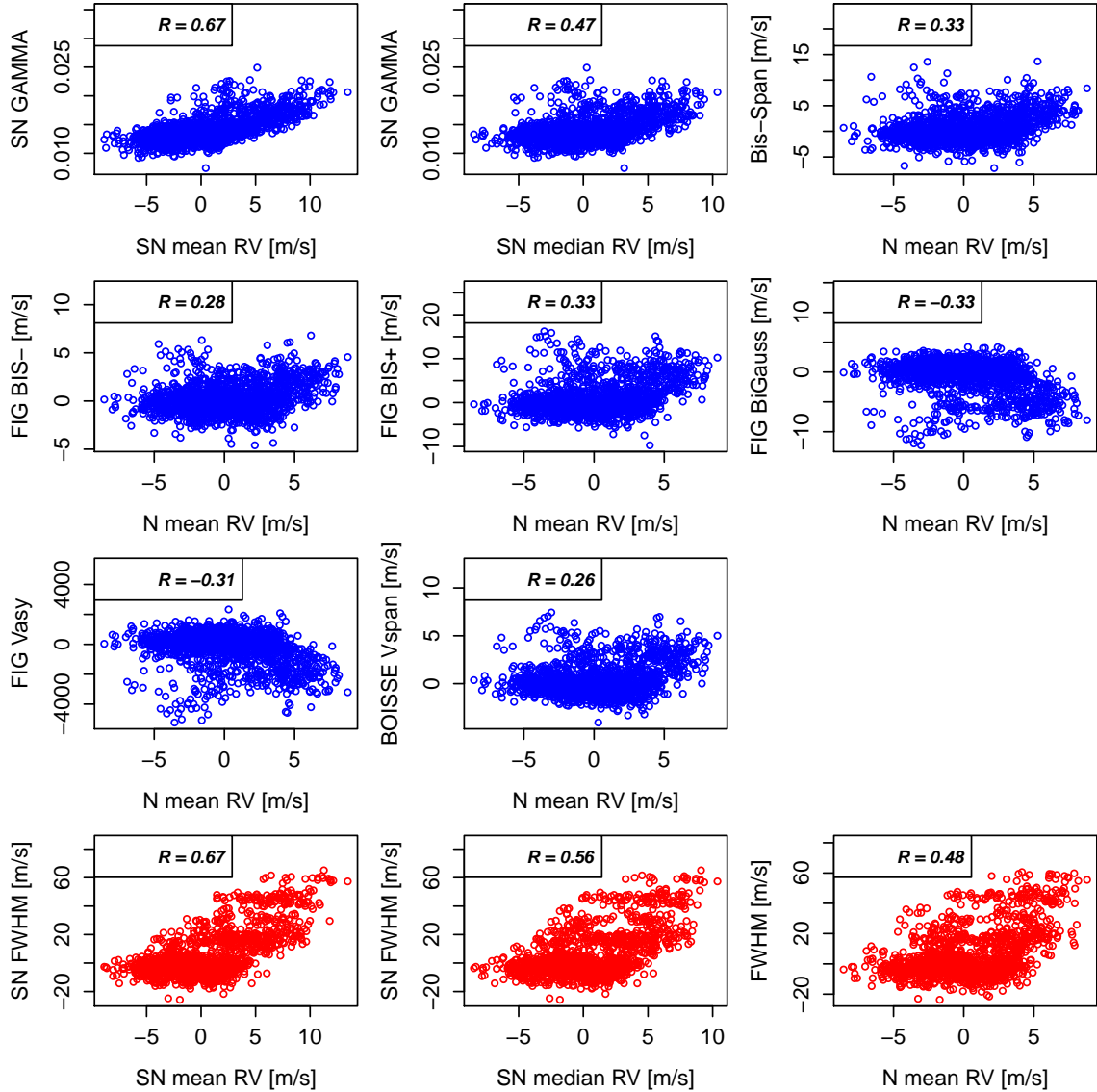| Star	R(Bis-Span, N mean RV)	R(FIG BiGaussian, N mean RV)	R(SN FWHM, SN mean RV)	R(FWHM, N mean RV)
HD192310	0.329(0.285; 0.373)	−0.333(−0.376; −0.289)	0.666(0.637; 0.692)	0.476(0.4367; 0.514)
HD10700	−0.073(−0.095; −0.0051)	0.127(0.105; 0.148)	0.421(0.403; 0.439)	0.529(0.513; 0.545)
HD215152	−0.067(−0.184; 0.052)	0.269(0.155; 0.376)	0.210(0.094; 0.321)	−0.138(−0.253; −0.020)
Corot 7	0.092(−0.058; 0.238)	−0.335(−0.228; −0.082)	−0.709(0.626; 0.776)	0.595(0.489; 0.683)

Fig. A.1. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD192310. (bottom row) Correlations between the FWHM and the estimated RVs. The correlations are consistently stronger when using parameters derived from the SN than the Normal. The estimated R are all statistically significant.

Fig. A.2. (top) The RVs (black dots) for HD192310 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1 \gamma + \beta_2 \text{SN FWHM}$ for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1 \text{BIS SPAN} + \beta_2 \text{FWHM}$ for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

Fig. A.3. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD10700. (bottom row) Correlations between the FWHM and the RVs for HD10700. The correlations are consistently stronger when using SN mean RV compared to N mean RV for the asymmetry parameters; however, the correlation between the FWHM and the N mean RV, only for this quiet star, is stronger the the analogous correlations with the estimated SN RVs. The estimated *R* are statistically significant, except for the correlation between FIG BIS and RV (p–values=0.36).

Fig. A.4. (top) The RVs (black dots) for HD10700 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\mathrm{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\mathrm{activity}} = \beta_0 + \beta_1$BIS SPAN$+\beta_2$FWHM for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

Fig. A.5. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD215152. (bottom row) Correlations between the FWHM and the RVs for HD215152. The correlations are consistently stronger when using SN mean RV compared to N mean RV. The p–values associated with each R are not statistically significant for the correlation between N mean RV and BIS SPAN (p–values=0.27), the correlation between N mean RV and FIG BIS- (p–values=0.05), the correlation between SN median RV and SN FWHM (p–values=0.5) and the correlation between N mean RV and FWHM (p–values=0.2).

Fig. A.6. (top) The RVs (black dots) for HD215152 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{activity} = \beta_0 + \beta_1\gamma + \beta_2 SN$ FWHM for the SN fit and on $RV_{activity} = \beta_0 + \beta_1$ BIS SPAN $+ \beta_2$ FWHM for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

Fig. A.7. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for Corot 7. (bottom row) Correlations between the FWHM and the RVs for Corot 7. The correlations are consistently stronger when using parameters derived from the SN than the Normal. The p–values associated with each R are not statistically significant for the correlation between N mean RV and BIS SPAN (p–values=0.23) and the correlation between N mean RV and FIG BIS- (p–values=0.11).

Fig. A.8. (top) The RVs (black dots) for Corot 7 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1 \gamma + \beta_2 \text{SN FWHM}$ for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1 \text{BIS SPAN} + \beta_2 \text{FWHM}$ for the normal fit. The residuals have a smaller systematic component when using the proposed model of Eq. 5 (black dots) compared to the usual model (blue triangles). Moreover, once corrected for stellar activity using Eq. 5, the remaining std from the SN models are 0.334 m s^{-1} smaller than the remaining std of the Normal model.